

**THEORETICAL FOUNDATIONS AND IMPLEMENTATION
OPPORTUNITIES OF ARTIFICIAL INTELLIGENCE AND COMPUTER
VISION TECHNOLOGIES IN AUTOMATING METROPOLITAN
ESCALATORS.**

Pardaboyev Humoyun Ravshan ogli

Basic Doctoral Student, “Tashkent State University of Economics” State Institution

ORCID: <https://orcid.org/0009-0003-3689-9820>

Email: h.pardaboyev@tsue.uz

Scientific Supervisor: Zaripov Bakhodir Bobomurod ogli

Tashkent State University of Economics, Department of “Artificial Intelligence”,

PhD in Economics, Associate Professor

Abstract. This study examines real-time monitoring and automatic management of passenger flow on metropolitan escalators using artificial intelligence and computer vision technologies. The aim is to develop the conceptual foundations of an automatic hazard detection system based on deep learning algorithms (CNN, YOLO, transformer). Systematic literature analysis based on PRISMA principles, comparative analysis, and conceptual modeling were employed, with 25 international sources analyzed. A three-layered system architecture integrating Edge AI and Fog Computing is proposed. The daily passenger flow of Tashkent Metropolitan exceeding one million in 2025 confirms the urgency of implementation.

Keywords: artificial intelligence, computer vision, deep learning, YOLO, metropolitan, escalator, Edge AI, video analytics, safety.

Introduction

The intensification of urbanization and the growth of megacity populations is significantly increasing the burden on urban transport, particularly on metropolitan

infrastructure. The Tashkent Metropolitan — the only metro network in Central Asia — saw its passenger flow grow from 38.8 million in 2020 to 290 million in 2025, i.e., more than sevenfold over five years [1, 2]. The daily passenger flow has exceeded 1 million, with the “Pakhtakor” station alone serving 77,000 passengers per day [3].

Under these conditions, the problem of congestion and accidents on escalators has become an urgent research topic. Traditional video surveillance systems rely on operator control, and due to the limitations of the human factor, real-time detection of hazardous situations is difficult [4]. Integrating artificial intelligence and computer vision technologies into metropolitan infrastructure is a promising solution [5, 6]. The aim of this study is to analyze the theoretical foundations of these technologies and to develop a conceptual model for their implementation in the conditions of the Tashkent Metropolitan.

Literature Review

The revolution in computer vision began with Krizhevsky et al.’s AlexNet convolutional model [7]. Subsequently, Vaswani et al. proposed the transformer architecture [8], which was later adapted for images by Dosovitskiy (ViT) [9]. Carion et al. developed DETR for object detection [10]. The YOLO family has evolved from YOLOv3 in 2018 [11] through YOLOv5, v7, v8, v9, and v11; Akkaya and Halici compared the accuracy and speed of the various versions [12], and a recent study confirmed the superiority of YOLOv7 and v11 in crowd detection [13].

In the metropolitan environment, Xu et al. developed the transformer-based MPDNet [14]. Wu et al. presented a bidirectional flow-counting system using improved DeepSORT at Shanghai and Shenzhen stations [15]. Pawar et al. performed real-time crowd assessment at stations based on YOLOv8s [16]. A distributed deep learning system demonstrated 94–99% accuracy [17]. Osipov et al. published an escalator safety system based on CNN and recurrent neural networks in the journal Nature [18]. Wang et al. analyzed the cascading spread of escalator accidents [19].

Palestinian researchers detected falls with 94.99% accuracy using the LRCN model [20]. The EN 115-1:2017 [21] and ISO 8102-6:2019 [22] standards regulate escalator safety at the normative level.

Research Methodology

A four-stage mixed methodology was applied in the study. In the first stage, a systematic literature review based on PRISMA principles was conducted: 25 Q1–Q2-level sources from 2019–2026 were selected from the IEEE Xplore, Springer Link, Elsevier, MDPI, Frontiers, Nature, and arXiv databases. In the second stage, a comparative analysis of YOLO and other models was carried out in terms of accuracy (mAP), speed (FPS), and resource requirements. In the third stage, official statistics of the Tashkent Metropolitan for 2020–2025 were studied. In the fourth stage, a conceptual model based on the Edge–Fog–Cloud architecture was developed.

Results and Discussion

Theoretical Foundations and Mathematical Modeling

Convolutional neural networks (CNN) automatically learn the spatial hierarchical features of an image. The convolution operation is expressed by the following formula:

$$(I * K)(i, j) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I(i+m, j+n) \cdot K(m, n) \quad (1)$$

where I is the input image, K is the kernel (filter), and M and N are the dimensions of the kernel. In the YOLO algorithm, the IoU (Intersection over Union) metric is used to evaluate object detection accuracy:

$$IoU = \frac{Area(B_{pred} \cap B_{gt})}{Area(B_{pred} \cup B_{gt})} \quad (2)$$

where B_{pred} is the bounding box predicted by the model and B_{gt} is the ground-truth bounding box. Detection is considered correct when $IoU \geq 0.5$. Crowd density and the risk index on the escalator are determined as follows:

$$\rho(t) = \frac{N(t)}{A_{escalator}}, \quad R_{risk} = \alpha \cdot \rho(t) + \beta \cdot v(t) \quad (3)$$

where $\rho(t)$ is the density at time t , $N(t)$ is the number of passengers, A is the escalator surface area, $v(t)$ is the speed of movement, and α and β are weighting coefficients. The total latency for real-time monitoring is calculated as follows:

$$T_{latency} = T_{capture} + T_{inference} + T_{transfer} + T_{decision} \quad (4)$$

For real-time monitoring, $T_{latency}$ must be ≤ 200 ms; otherwise, timely detection of a fallen passenger becomes impossible.

Comparative Analysis of YOLO Models

Model	Accuracy (mAP)	Speed (FPS)	Model Size	Main Advantage
YOLOv5s	56.8%	125	14 MB	Highest speed, suitable for Edge devices
YOLOv7	69.7%	96	36 MB	High performance in crowd detection
YOLOv8	70.3%	110	22 MB	Balance of accuracy and speed

Model	Accuracy (mAP)	Speed (FPS)	Model Size	Main Advantage
YOLOv11	73.5%	105	19 MB	Most advanced, efficient compact module

Table 1. Comparative technical analysis of YOLO models

As shown in Table 1, the YOLOv8 and YOLOv11 models provide the best balance of accuracy and speed for real-time monitoring.

4.3. Passenger Flow of the Tashkent Metropolitan

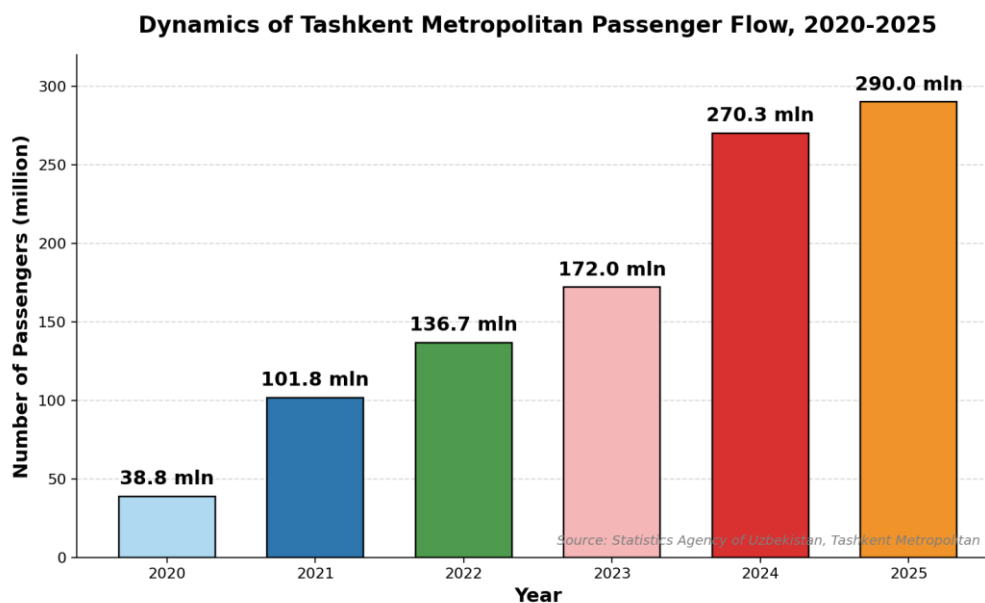


Figure 1. Dynamics of Tashkent Metropolitan passenger flow, 2020–2025.

No.	Station Name	Daily Passengers	Direction
1	Pakhtakor	77,272	Chilanzar
2	Amir Temur Avenue	60,048	Yunusabad

No.	Station Name	Daily Passengers	Direction
3	Chilanzar	58,754	Chilanzar
4	Buyuk Ipak Yoli (Great Silk Road)	54,195	Uzbekiston

Table 2. Most heavily loaded stations of the Tashkent Metropolitan (October 2024).

Given such high loads, it is advisable to begin the pilot system precisely at these stations.

Proposed System Architecture

Based on the analysis results, a three-layer intelligent monitoring system architecture is proposed (Figure 2). The Edge layer includes local computing devices near the escalator (NVIDIA Jetson, Google Coral), equipped with YOLOv8n or YOLOv5s models, processing video streams with 50–100 ms latency. The Fog layer consists of station-level servers that aggregate data and make autonomous decisions. The Cloud layer is located in a central data center and performs long-term analysis and model retraining tasks.

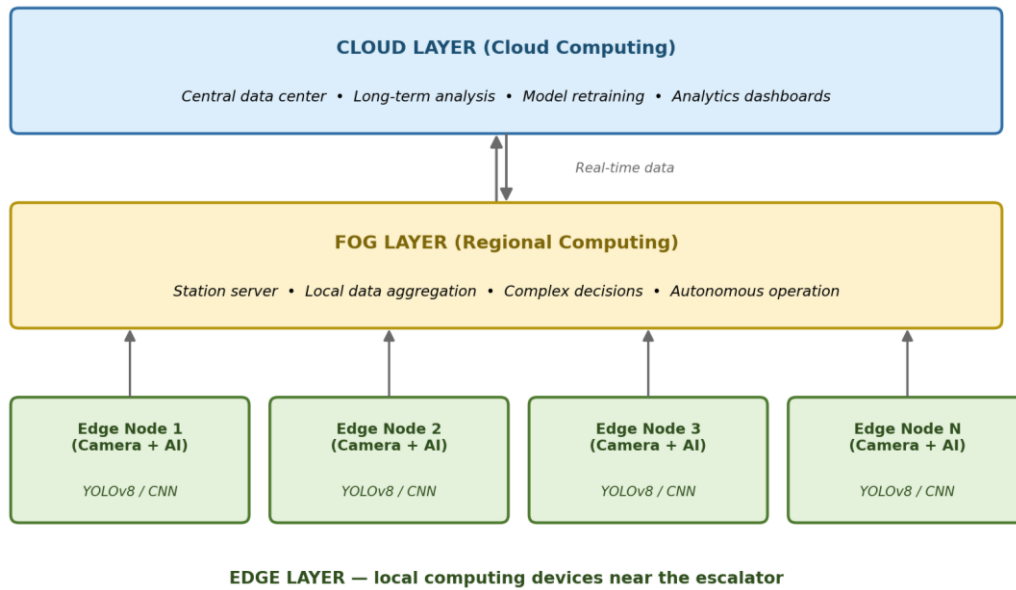


Figure 2. Three-layer architecture of the escalator monitoring system.

Hazardous Situations on Escalators and Detection Methods

Hazardous Situation	Detection Method	Accuracy	Source
Fall	Pose estimation (OpenPose, HRNet)	94.99%	[20, 23]
Crowd congestion	Density map regression	92-95%	[16, 17]
Walking against direction	Optical flow + CNN	90-93%	[18]
Anomalous behavior	Autoencoder + LSTM	88-92%	[18]

Table 3. Hazardous situations on escalators and artificial intelligence methods.

The advantage of the pose estimation approach is that, since it operates on a skeletal basis, it requires fewer computational resources and preserves privacy [23].

Implementation Opportunities in the Tashkent Metropolitan

A phased implementation of the system is proposed: Phase 1 (6–12 months) — pilot project at the Pakhtakor and Khalqlar Do'stligi stations; Phase 2 (12–24 months) — expansion to all stations of the Chilanzar line; Phase 3 (24–36 months) — full integration across all 50 stations and connection with the city management system. From the standpoint of personal data protection, the skeleton-based pose estimation approach is preferable [23].

Conclusion

The study shows that artificial intelligence and computer vision technologies — CNN, YOLO, transformer, and pose estimation methods — provide high efficiency in automating metropolitan escalators. The integration of Edge AI, Fog Computing, and Cloud Computing technologies ensures high response speed, reliability, and scalability. For the implementation of an intelligent monitoring system in the conditions of the Tashkent Metropolitan, the following is recommended: (1) launch a pilot project at the most heavily loaded stations; (2) create a local dataset and retrain models; (3) align with international standards (EN 115, ISO 8102-6); (4) comply with personal data protection norms. Future research should be directed toward creating a dedicated dataset for Tashkent Metropolitan escalators and developing hybrid model architectures.

References

1. The number of passengers transported on the Tashkent metro increased by 20 million in 2025 // Gazeta.uz. — 2026.01.13. — URL: <https://www.gazeta.uz/oz/2026/01/13/metro/>
2. Passenger flow on the Tashkent metro increased by 57% in 2024 // Gazeta.uz. — 2025.01.27.

3. How does the Tashkent Metropolitan plan to solve the congestion problem? // Gazeta.uz. — 2024.10.14.
4. Sari M.I., Ibnugraha P.D. et al. Initial Estimation Method for Public Transport Passenger Density // IEEE Proc. — 2023. — DOI: 10.1109/GCAT59970.2023.10353287.
5. Distributed Deep Learning System for Real-Time Passenger Flow Measurement // Applied Intelligence (Springer). — 2025. — DOI: 10.1007/s10489-025-06954-9.
6. AI-Powered Safety Monitoring on Metro Platforms with Computer Vision // J. Propulsion Technology. — 2025. — Vol. 46, No. 2.
7. Krizhevsky A., Sutskever I., Hinton G.E. ImageNet Classification with Deep CNN // Comm. of the ACM. — 2017. — Vol. 60, No. 6. — P. 84–90.
8. Vaswani A. et al. Attention Is All You Need // NeurIPS. — 2017. — Vol. 30.
9. Dosovitskiy A. et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale // ICLR. — 2021.
10. Carion N. et al. End-to-End Object Detection with Transformers // ECCV. — 2020. — P. 213–229.
11. Redmon J., Farhadi A. YOLOv3: An Incremental Improvement // arXiv. — 2018. — arXiv:1804.02767.
12. Akkaya I.B., Halici U. A YOLO-based method for real-time crowd detection // J. Real-Time Image Processing. — 2023. — Vol. 20. — DOI: 10.1007/s11554-023-01290-y.
13. Evaluating YOLO Models for Detecting Crowds in Sparse Regions // Springer LN. — 2025. — DOI: 10.1007/978-981-96-8796-1_5.
14. Xu D. et al. MPDNet: A Transformer-based real-time passenger detection network // Frontiers in Physics. — 2022. — Vol. 10. — DOI: 10.3389/fphy.2022.1017951.

15. Wu J. et al. Bi-directional passenger flow tracking based on improved Deep-Sort // Measurement and Control. — 2024. — Vol. 57, No. 1.
16. Pawar Y. et al. Metro Crowd Management using YOLOv8s // SSRN. — 2025. — DOI: 10.2139/ssrn.5397104.
17. Real-Time Passenger Counting Using YOLO-Based Deep Learning // J. Machine Learning and Signal Processing. — 2025.
18. Osipov V. et al. Intelligent escalator passenger safety management // Scientific Reports (Nature). — 2022. — Vol. 12. — DOI: 10.1038/s41598-022-09498-x.
19. Wang Z. et al. Critical hazards identification and prevention of cascading escalator accidents // IJERPH. — 2020. — Vol. 17. — DOI: 10.3390/ijerph17103400.
20. Escalator Safety System Using Computer Vision and Deep Learning Based on LRCN // Palestine Polytechnic Univ. — 2024.
21. EN 115-1:2017. Safety of escalators and moving walks — Part 1: Construction and installation. — CEN, 2017.
22. ISO 8102-6:2019. Programmable electronic systems for escalators (PESSRAE). — ISO, 2019.
23. Closed-Circuit Television Data as an Emergent Data Source for Urban Rail Platform Crowding Estimation // arXiv. — 2025. — arXiv:2508.03749.
24. Research Into Real-Time Video Monitoring of Escalators // Elevator World. — 2022.
25. Jiang X.X.M., Niu S.S.G., Guo L.L.Z. Braking distance monitoring system for escalator // Advanced Manufacturing and Automation (Springer). — 2018. — P. 197–205.